NEON USER GUIDE TO STREAM DISCHARGE RATING CURVE (NEON.DP4.00133)

PREPARED BY	ORGANIZATION	DATE
Kaelin M Cawley	AQU	04/12/2018
Nick Harrison	AQU	04/12/2018

CHANGE RECORD

REVISION	DATE	DESCRIPTION OF CHANGE
A	01/11/2018	Initial Release

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

1 DESCRIPTION

1.1 Purpose

This document provides an overview of the data included in this NEON Level 4 data product, which is generated from Level 1 OS data, and associated metadata. In the NEON data products framework, the raw data collected in the field (i.e. staff gauge and discharge measurements from a single collection event) are considered the lowest level (Level 0). Raw data that have been quality checked and simple metrics that emerge from the raw data are considered Level 1 data products. Level 4 data products rely on inputs of any level data, often L1 or higher, and may involve calculations that use data collected over a range of spatial or temporal scales.

The text herein provides a discussion of measurement theory and implementation, data product provenance, quality assurance and control methods used, and approximations and/or assumptions made during L4 data creation.

1.2 Scope

This document describes the steps needed to generate the L4 data product Stream Discharge Rating Curve (NEON.DP4.00133) - the parameters that describe the relationship between staff gauge measurements and discharge measurements made at a stream over a range of dates and flow conditions - and associated metadata from input data and calculations. This document also provides details relevant to the publication of the data products via the NEON data portal, with additional detail available in the file, NEON Data Variables for Stream Discharge Rating Curve (NEON.DP4.00133) (AD[03]), provided in the download package for this data product.

This document describes the process for performing custom calculations derived from L1 Stream discharge field collection data (NEON.DP1.20048), geolocation data, and previous L4 Stream discharge rating curve data. For information of the raw data that are ingested and processed for the source data product see NEON.DP0.20048.001_dataValidation.csv and NEON User Guide to Stream Discharge (NEON.DP1.20048) available for download with the L1 data package. Please note that raw or lower level source data products (denoted by ‘DP0’) may not always have the same numbers (e.g., ‘20048’) as the corresponding L1 or L4 data product.

2 RELATED DOCUMENTS AND ACRONYMS

2.1 Associated Documents

AD[01]	NEON.DOC.000001	NEON Observatory Design (NOD) Requirements
AD[02]	NEON.DOC.002652	NEON Level 1, Level 2 and Level 3 Data Products Catalog
AD[03]	NEON.DP4.00133.001 _variables.csv	NEON Data Variables for Stream Discharge Rating Curve (NEON.DP4.00133)
AD[04]	NEON.DOC.001152	NEON Aquatic Sampling Strategy
AD[05]	NEON.DOC.000008	NEON Acronym List
AD[06]	NEON.DOC.000243	NEON Glossary of Terms

2.2 Acronyms

Acronym	Definition
Q	Stream discharge

3 DATA PRODUCT DESCRIPTION

The stream discharge rating curve data product provides a stage-discharge rating curve that describes the relationship between staff gauge and stream discharge measurements at stream and river sites each water year (October 1st - September 30th). The relationship developed for the stage-discharge rating curve is used for computing Continuous Stream Discharge (NEON.DP4.00133). An example rating curve that was plotted for a range of discharge (y-axis) and stage values (x-axis) is plotted below (Figure 1).

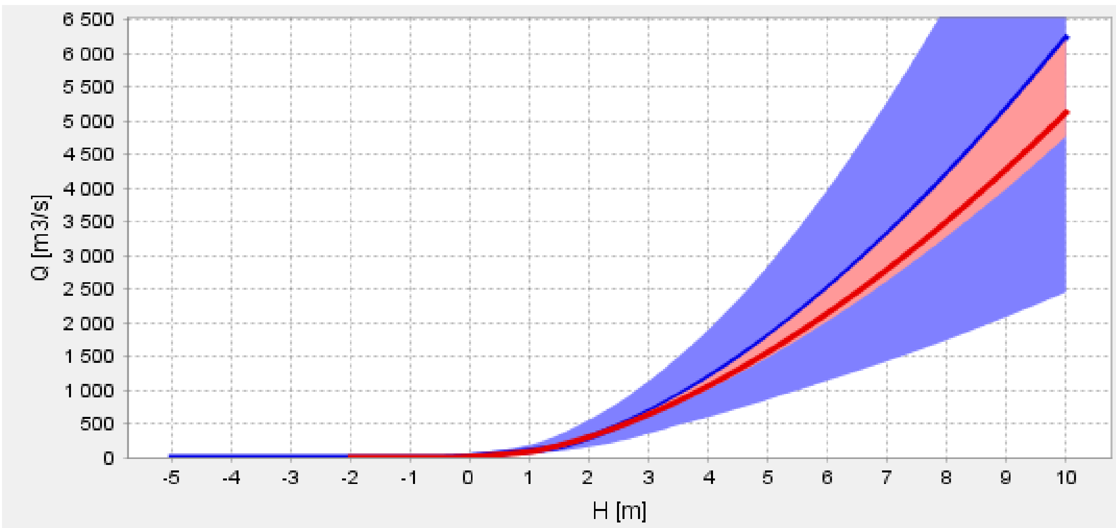

Figure 1: An example rating curve (red line) with parametric uncertainty (light red shading). Blue line and shading represent "prior" rating curve and parametric uncertainty using information based on hydraulic controls only, respectively. Picture taken from "Rating Curve" documentation for BaRatinAGE (Le Coz, 2013; Le Coz, 2014).

3.1 Spatial Sampling Design

The rating curve is developed at the site level using data collected at the same spatial scale as the discharge and stream stage measurements, which are measured at all wadeable stream sites on a run or riffle near sensor set #1 or sensor set #2 and near the buoy at river sites. Point measurements of water depth and velocity are made along the transect from one bank to the other using a wading rod and attached velocity meter (at wadeable stream sites) or acoustic Doppler current profiler instrumentation (at non-wadeable river sites). If a discharge cross-section or staff gauge is re-located, a new rating curve may be developed for a site if the hydrologic data is no longer directly comparable to the existing set of stage and discharge measurements.

3.2 Temporal Sampling Design

Stream discharge rating curve(s) are created or updated following the conclusion of the water year (October 1st to September 30th).

Table 1: List of input data for Stream Discharge Rating Curve.

Data Source	**DPID**	**table**
L1 Stream discharge field collection data	NEON.DP1.20048	dsc_fieldData
L1 Stream discharge field collection data	NEON.DP1.20048	dsc_individualFieldData
L4 Stream Morphology Map	NEON.DP4.00131	geo_controlInfo
L4 Stream Morphology Map	NEON.DP4.00131	geo_priorParameters

Table 2: List of geolocation inputs for Stream Discharge Rating Curve.

Data Source	**field**	**location**
NEON geolocation database	location startDate	SITE.AOS.gauge
NEON geolocation database	location endDate	SITE.AOS.gauge
NEON geolocation database	location zOffset	SITE.AOS.gauge

3.3 Variables Reported

All data and geolocation variables used as inputs for the development of a stream discharge rating curve are listed in Table 1 and Table 2. All variables reported in the published data (L4 data) are also provided separately in the file, NEON Data Variables for Stream Discharge Rating Curve (NEON.DP4.00133) (AD[03]).

Field names have been standardized with Darwin Core terms (http://rs.tdwg.org/dwc/; accessed 16 February 2014), the Global Biodiversity Information Facility vocabularies (http://rs.gbif.org/vocabulary/gbif/; accessed 16 February 2014), the VegCore data dictionary (https://projects.nceas.ucsb.edu/nceas/projects/bien/wiki/VegCore; accessed 16 February 2014), where applicable. For AOS, Earth Gravitational Model 96 (EGM96) is the reference gravitational ellipsoid. Latitudes and longitudes are denoted in decimal notation to six decimal places, with longitudes indicated as negative west of the Greenwich meridian.

Some variables described in this document may be for NEON internal use only and will not appear in downloaded data.

3.4 Spatial Resolution and Extent

The finest spatial resolution at which spatial data are reported is a site.

3.5 Temporal Resolution and Extent

The finest temporal resolution at which temporal data are reported is the date and time of a discharge and stage measurement for the individual points used to compute a rating curve for the sdrc_gaugeDischargeMeas table. The finest temporal resolution at which temporal data are reported is the water year for the rating curve specific tables: sdrc_stageDischargeCurveInfo, sdrc_sampledParameters, sdrc_posteriorParameters, sdrc_resultsResiduals.

3.6 Associated Data Streams

The data from this L4 data product is derived from a level 1 (L1) data product: Stream discharge field collection (NEON.DP1.20048). These data products can be linked by **siteID**.

The data from this L4 data product is derived from a level 4 (L4) data product: Stream Morphology Map data product (NEON.DP4.00131). These data products can be linked by **siteID**.

The data from this L4 data product is used to develop a level 4 (L4) data product: Continuous Stream Discharge (NEON.DP4.00130). These data products can be linked by **siteID**.

3.7 Product Instances

The NEON Observatory contains 24 wadeable streams and 3 large rivers.

Stream Discharge rating curve yields approximately 27 ratings curves (1 per site) per water year comprised of approximately 702 stage and discharge pair records per year during stage-discharge relationship development and 324 stage and discharge records per year once a relationship has been developed.

3.8 Data Relationships

The algorithm used for this L4 data product produces as many records in sdrc_gaugeDischargeMeas as there are for a site for the past water year in the dsc_fieldData table for Stream discharge field collection (NEON.DP1.20048). The values may differ in the two tables, though. Discharge values are recalculated from point measurements and stage values are offset using information from the NEON geolocation database. One record is created for each hydraulic control at a site for the sdrc_posteriorParameters table. One record is created for each rating curve (usually one per year) for the sdrc_stageDischargeCurveInfo table. 500 records per hydraulic control at a site are created for the sdrc_sampledParameters, which represent the Markov Chain Monte Carlo (MCMC) samples of the range of model parameters. One record per staff gauge/discharge measurement set used to develop the rating curve is created in the sdrc_resultsResiduals table, this is often more than the number of records in the sdrc_gaugeDischargeMeas due to the inclusion of measurements from the previous water year in the development of the curve.

sdrc_gaugeDischargeMeas.csv - > One record expected per staff gauge and discharge measurement collected for a site in the past water year

sdrc_posteriorParameters.csv - > One record expected per hydraulic control at a site

sdrc_stageDischargeCurveInfo.csv - > One record expected per annual rating curve developed for a site

sdrc_sampledParameters.csv - > 500 records expected per hydraulic control at a site sdrc_resultsResiduals.csv - > One record expected per staff gauge and discharge measurement used to develop the rating curve for the current water year

geo_controlInfo.csv - > One record per the square of the number of hydraulic control(s) at a site, e.g. there will be 4 records if there are 2 controls. This table is an exact copy of the table of the same name in the download package for the Stream Morphology Map data product (NEON.DP4.00131) and is include here for its value in evaluating Equation 1.

gaugeEventID can be used to link data in the sdrc_gaugeDischargeMeas and sdrc_resultsResiduals tables. There may be multiple entries for a given **gaugeEventID** in the sdrc_resultsResiduals table if the staff gauge and discharge readings were used in more than one rating curve, which is common if the stream or river does not undergo major geomorphic change between water years.

allEventID contains a concatenated string, which can be parsed using a "|" delimiter to determine which **gaugeEventIDs** were used for developing a specific rating curve. Parsing **allEventID** will enabe linking the sdrc_stageDischargeCurveInfo table with the sdrc_gaugeDischargeMeas and sdrc_resultsResiduals tables.

curveID can be used to link the sdrc_posteriorParameters, sdrc_stageDischargeCurveInfo, sdrc_sampledParameters, and sdrc_resultsResiduals tables.

4 Algorithm Theoretical Basis

4.1 Theory of measurement

Stream discharge is the volume of water flowing through a stream over time and is a function of the height of the water column, cross-sectional area, and water velocity. In a natural stream, riffles can act as weirs and the stream velocity is often a function of channel roughness where natural weirs are not present. The stream channel's physical dimension controls the cross-sectional area as the water column height changes (Figure 2). Staff gauges are used to measure the surface water elevation relative to a fixed reference. The staff gauge does not measure absolute water depth. Staff gauge measurements are collected at the start and end of stream discharge measurements where wading surveys are utlized in wadeable streams and acoustic Doppler current profiler instrumentation (ADCP) are utilized in non-wadeable rivers. With these paired measurements and the physical hydraulic characteristics, a stage-discharge rating curve is developed using the Bayesian Modeling (BaM) algorithm, described below.

4.2 Theory of Algorithm

The NEON stage-discharge rating curve is developed using a Bayesian modeling technique developed by the Bayesian Rating Curve Advanced Graphical Environment (BaRatinAGE) development team (Le Coz et al., 2013; Le Coz et al., 2014). The executable and/or a GUI is available freely with an individual license by sending an email to: baratin.dev@lists.irstea.fr.

The rating curve relies on a "prior" rating curve that is developed for the hydraulic controls. The physical dimensions of the channel, the number of hydraulic controls selected, and the physical dimensions of the hydraulic con-

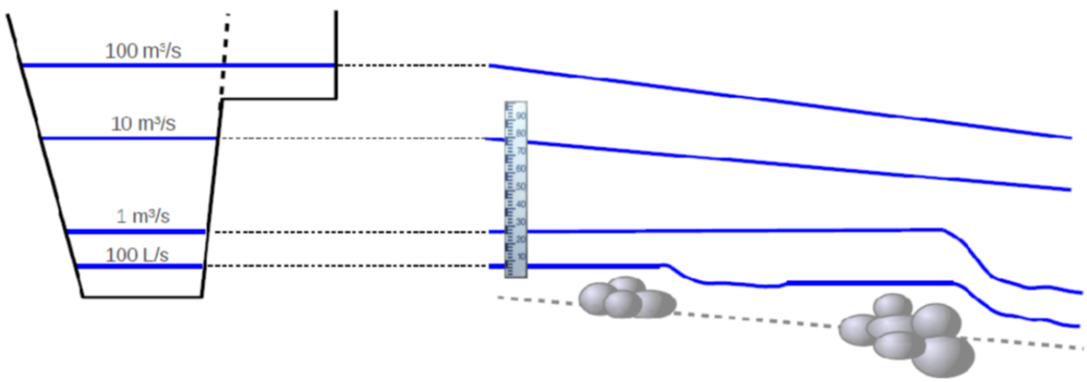

Figure 2: Diagram of natural riffles acting like weirs at low water and channel controled at higher water level. Picture taken from "Hydraulic Analysis" documentation for BaRatinAGE (Le Coz et al., 2013; Le Coz et al., 2014).

trols are derived from cross-section survey data. Exponential equations for each control are then calculated (Equation 1). A "posterior" rating curve is then fit using the "prior" rating curve and the gauging records using Bayesian estimation of the rating curve and a Markov Chain Monte Carlo (MCMC) sampling (Le Coz, 2014). Equation 1 and text below are taken from "Rating curve equation" documentation for BaRatinAGE (Le Coz et al., 2013).

$$Q(h) = \sum_{r=1}^{N_{segment}} \left(1_{[K_{r-1};K_r]}(h) \times \sum_{j=1}^{N_{control}} M(r,j) \times a_j (h - b_j)^{c_j} \right) \tag{1}$$

In the above equation, $M(r, j)$ is the matrix of controls, and the notation $1_I(h)$ denotes a function equal to 1 if h is included in the interval I, and zero otherwise. This equation shows that the stage discharge relation is a combination of power functions, and the matrix of controls is used to specify how this combination operates (succession or addition of controls).

Users are directed to the information in the **geo_controlInfo.csv** table for the data needed for the $1_I(h)$ function.

4.3 Algorithm Implementation

The NEON OS transition system runs a Docker container containing R code to create a rating curve every 366 days (to account for leap years) to produce the annual rating curve (similar infrastructure to Metzger, 2017 without the use of HDF5 file formats).

Within the Docker container:

1. Data for the site and water year is queried from the NEON database.
2. Configuration files for BaM are created.
3. BaM executable is run.
4. BaM outputs are written to the L4 data tables in the NEON database for publication to users.

5 DATA QUALITY

5.1 Data Entry Constraint and Validation

Many quality control measures are implemented at the point of data entry (i.e., the L1 data that is used as an input for this data product) within a mobile data entry application or web user interface (UI).

5.2 Data Revision

All data are provisional until a numbered version is released; the first release of a static version of NEON data, annotated with a globally unique identifier, is planned to take place in 2020. During the provisional period, QA/QC is an active process, as opposed to a discrete activity performed once, and records are updated on a rolling basis as a result of scheduled tests or feedback from data users. The Change Log section of the data product readme, provided with every data download, contains a history of major known errors and revisions.

5.3 Uncertainty

One of the benefits of using BaM and MCMC sampling is that there are a large number of realizations from the posterior distribution, which can be used to quantify uncertainty associated with the maximum liklihood posterior parameters (BaRatin statistical model documentation and Le Coz et al., 2014). NEON publishes both the parametric and remnant (structural) error based off of 500 realizations from the posterior distribution.

Note that the uncertainty published in the NEON data downloads is expanded uncertainty, i.e. multiplied by a factor of 1.96 to cover two standard deviations. When using the BaRatin GUI tool the uncertainty should be represented the same was as NEON publishes it. For the BaM executable, though, uncertainty is represented as one standard deviation. So, the NEON data should be divided by a factor of 1.96 before writing out data and configurations.

5.4 Quality Flagging

This data product contains a **dataQF** field in each data record that is a quality flag for known errors applying to the record. Please see below for an explanation of **dataQF** codes specific to this product.

In addition to the **dataQF** field in each table there is also a **recalculatedL1QF** that indicates if the recalculated discharge value is different than the L1 published discharge value (1 indicates they are different and 0 indicates that they are not different) and a **L1DataQF** that indicates a quality flag associated with the L1 input data.

fieldName	value	definition
dataQF	legacyData	Data recorded using a paper-based workflow that did not implement the full suite of quality control features associated with the interactive digital workflow

6 REFERENCES

Le Coz, J., B. Renard, L. Bonnifait, F. Branger, R. Le Boursicaud (2014) Combining hydraulic knowledge and uncertain gaugings in teh estimation of hydrometric rating curves: A Bayesian approach. *Journal of Hydrology*, 509:573-587. DOI: 10.1016/j.hydrol.2013.11.2016

Le Coz, J., C. Chaleon, L. Bonnifait, R. Le Boursicaud, B. Renard, F. Branger, J. Diribarne, M. Valente (2013) Bayesian analysis of rating curves and their uncertainties: the BaRatin method. *Houille Blanche-Revue Internationale De L Eau*, 6:31:41. DOI: 10.1051/lhb/2013048.

Metzger, S., D. Durden, C. Sturtevant, H. Luo, N. Pingintha-Durden, T. Sachs, A. Serafimovich, J. Hartmann, J. Li, K. Xu, A. Desai (2017) eddy4R: A community-extensible processing, analysis and modeling framework for eddy-covariance data based on R, Git, Docker and HDF5. *Geoscientific Model Development Discussions*, 1-26. DOI: 10.5194/gmd-2016-318.